March 31, 2025

Hon. Ona T. Wang
United States Magistrate Judge
Southern District of New York

> Re: *The New York Times Company v. Microsoft Corporation, et al.*, 1:23-cv-11195; *Daily News, LP, et al. v. Microsoft Corporation, et al.*, 1:24-cv-3285; *The Center for Investigative Reporting, Inc. v. OpenAI, Inc.*, No. 1:24-cv-04872

Dear Magistrate Judge Wang:

I write on behalf of Plaintiffs Daily News, LP, *et al.* (the "*Daily News* Plaintiffs), The New York Times Company ("The Times") and the Center for Investigative Reporting ("CIR") (collectively, the "News Plaintiffs"). We ask that you order Microsoft to: (1) provide discovery into all versions of Microsoft's Copilot chatbot product; and (2) provide News Plaintiffs access to inspect the "Deucalion" large language model ("LLM"), which Microsoft fined-tuned at OpenAI and deployed in its commercial Copilot product.

1.   DISCOVERY INTO MICROSOFT'S COPILOT PRODUCT

News Plaintiffs allege in their Complaints that Microsoft's Copilot chatbot infringes News Plaintiffs' copyrights because, *inter alia*, Copilot is powered by: (i) the GPT family of LLMs trained on News Plaintiffs' content, and (ii) Microsoft's Bing index that provides copies of News Plaintiffs' content to those LLMs in order to create "synthetic search results." *See, e.g.*, *The New York Times* Original Compl. (Dkt. 1) ¶ 2 ("Through Microsoft's Bing Chat (recently rebranded as 'Copilot') and OpenAI's ChatGPT, Defendants seek to free-ride on The Times's massive investment in its journalism by using it to build substitutive products without permission or payment."), *id.* at ¶¶ 83, 92, 98, 102, 108-117, 166-167 (alleging ongoing infringement); *Daily News* Compl. (Dkt. 1) ¶¶ 1, 80, 90, 96, 114-138, 199-200 (alleging ongoing infringement). Although Microsoft has continuously offered the Copilot product to the public under the same name since the filing of the Complaints, during a February 10, 2025, call among News Plaintiffs' technical experts and Microsoft's outside consultants to discuss Microsoft's output logs, Microsoft's consultants revealed for the first time that the output log data under discussion only relates to a version of Copilot that Microsoft deprecated on October 1, 2024 ("deprecated Copilot") and not the "Copilot" product that is currently in use.

In a follow up meet and confer on February 27, Microsoft argued that it was justified in withholding discovery into what it characterized as the "refreshed" Copilot launched on October 1, 2024, because that version of Copilot is purportedly a different product than the "Copilot" named in the Complaints. Maisel Decl.¶ 2, Ex. 1 (blog post dated Oct. 1, 2024). More specifically, although "refreshed" Copilot has the same name as deprecated Copilot, is also powered by the GPT LLMs, and also generates synthetic search results using the Bing index just like deprecated Copilot, Microsoft contends that "refreshed" Copilot is outside the scope of the Copilot product identified in the Complaints because it has a different software architecture and implicates different documents and custodians.

But nothing in the Complaints or discovery to date justifies the line Microsoft has drawn

March 31, 2025
Page 2

between deprecated versions of Copilot (concededly relevant) and refreshed Copilot. Plaintiffs' allegations relate to "Copilot" without date restriction. Both versions of Copilot are chatbots powered by GPT LLMs using the Bing index (and News Plaintiffs' content). And both Copilots infringe News Plaintiffs' copyrights in the same way. Following Microsoft's revelation that it was withholding discovery into the "refreshed" Copilot, News Plaintiffs confirmed that "refreshed" Copilot–just like its same-named predecessor–also regurgitates News Plaintiffs' content. Maisel Decl., ¶¶ 3-4, Exs. 2-3. In short, the "refreshed" Copilot (just like the "deprecated" Copilot before it) uses an OpenAI LLM that appears to have been trained on News Plaintiffs' content, has memorized that content, can regurgitate that content to users when prompted, and copies News Plaintiffs' content through retrieval augmented generation when generating synthetic search results.

Microsoft's allegations that the "refreshed" Copilot is built on a different architecture and has different documents and custodians that are relevant to it only bolsters arguments that discovery into it is relevant and non-duplicative.[1] News Plaintiffs' allegations regarding Microsoft's Copilot product are not limited to a particular architecture, documents, or custodians (which were unknown to News Plaintiffs at the time). Rather, the allegations relate to the manner in which the underlying LLMs are trained, their propensity to memorize and regurgitate the content on which they were trained—including News Plaintiffs' content—and their use in generating synthetic search results using retrieval augmented generation.

Microsoft's position that the Court's prior order on future, anticipated, or later products (Dkt. 365 at 5) shields discovery into updated versions of Copilot is also not tenable. The Court made clear during the oral hearing on the motion that the "[r]uling is that [discovery is] limited to Microsoft products named in the complaint." Transcript of December 3, 2024 Hearing, p. 90. Microsoft's Copilot product, regardless of whether Microsoft now characterizes a particular version of that product as being "deprecated" or "refreshed," is named in the News Plaintiffs' complaints.[2] Accordingly, discovery into the "refreshed" Copilot is plainly relevant and within scope of the claims in this case, and should be produced. *See, e.g.*, *Solid Oak Sketches, LLC v. 2K Games, Inc.*, No. 16CV724-LTS, 2016 WL 4126543, at *3 (S.D.N.Y. Aug. 2, 2016) (holding that "updates to the title and visual and graphical improvements" of a video game did not transform the updated version of that game into a different product for determining when copyright infringement began).

News Plaintiffs request that the Court order Microsoft to provide discovery into all of Copilot, including its "refreshed" version. This discovery should include Microsoft (1) updating

---

[1] Microsoft has not otherwise explained how the refreshed and deprecated Copilot products differ from each other. Maisel Decl., ¶ 5.

[2] *See, e.g.*, *Daily News* Compl., ¶¶ 1 (referring to "Microsoft Copilot (formerly known as Bing Chat)"), 96 (alleging copyright infringement by "Copilot, Copilot Pro, and the Microsoft 365 Copilot line of digital assistants"); *see also The New York Times* Compl., ¶¶ 2 (referring to "Bing Chat (recently rebranded as 'Copilot')"), 102 (alleging copyright infringement by "Bing Chat, and the Microsoft 365 Copilot line of digital assistants").

March 31, 2025
Page 3

its responses to News Plaintiffs' interrogatories to encompass information about the refreshed Copilot; (2) searching for and producing documents responsive to News Plaintiffs' requests for production that relate to the refreshed Copilot, including producing documents that were withheld on the basis that they relate to the refreshed Copilot; and (3) producing output log data, source code, training data, and any models for the refreshed Copilot in response to News Plaintiffs' requests for inspection.

2.   ACCESS TO MICROSOFT'S DEUCALION MODEL

New Plaintiffs inspected OpenAI's commercial LLMs consistent with the terms ordered by this Court. *See* Dkt. 365 at 3 (directing OpenAI to provide News Plaintiffs with one week of free searches). That inspection, however, did not include one bespoke LLM called "Deucalion," which is based off an OpenAI GPT model that Microsoft fine-tuned at OpenAI and commercially deployed in the deprecated version of its Copilot product (formerly known as Bing Chat). Similar to News Plaintiffs' inspection of OpenAI's commercial models, News Plaintiffs now seek to inspect Microsoft's Deucalion model, consistent with their requests for inspection that have been served on Microsoft. Microsoft has refused on grounds that Deucalion is not relevant. The refusal is without merit.

Access to Microsoft's Deucalion model will allow News Plaintiffs to determine the amount of their content that the model has "memorized," which is relevant to News Plaintiffs' allegations that the LLMs regurgitate and/or encode infringing copies or derivatives of News Plaintiffs' content upon which the LLMs were trained. *See, e.g.*, *Daily News* Compl. (Dkt. 1), ¶ 195 (alleging that "by storing, processing, and reproducing the GPT models … Microsoft and the OpenAI Defendants have jointly directly infringed the Publishers' exclusive rights in their copyrighted works"); *see also id.*, ¶¶ 77 (alleging that memorization "shows that LLM parameters encode retrievable copies of many of those training works"), 96, 192 (alleging that the asserted works "have been distributed by and encoded within Defendants' GPT models"); *see also The New York Times* Compl. (Dkt. 1), ¶¶ 80, 98, 159, 162. Memorization is also relevant to Defendants' anticipated fair use defense, including the argument that the LLM training process is a transformative use of News Plaintiffs' copyrights.

Microsoft has not articulated an undue burden in making Deucalion available for inspection to News Plaintiffs. It could not do so. Instead, Microsoft argues that access to Deucalion is not relevant because end users never had "direct access" to the model because such access was mediated through other software layers with the Copilot product. But irrespective of end user's direct access, the model itself is relevant to News Plaintiffs allegations concerning model memorization, including: (1) the model itself is infringing and not transformative (*e.g.*, because it encodes copies of the asserted works), and (2) memorized (or regurgitated) outputs are further evidence of the extent and nature of Defendants' unauthorized use of News Plaintiffs' content during training and as part of Copilot's outputs. Because the information sought is relevant and proportional to the claims in this case, News Plaintiffs ask the Court to order Microsoft to provide News Plaintiffs access to inspect the Deucalion model.

March 31, 2025
Page 4

March 31, 2025   Respectfully submitted,

                                                 /s/ Steven Lieberman
                                                 Steven Lieberman
                                                 Rothwell, Figg, Ernst & Manbeck, P.C.

                                                 /s/ Ian Crosby
                                                 Ian B. Crosby
                                                 Susman Godfrey L.L.P.

                                                 /s/ Stephen Match
                                                 Stephen Match
                                                 Loevy & Loevy

    cc: All Counsel of Record (via ECF)